Jurnal Varian
Vol. 6, No. 1, Oktober 2022, Hal. 1–10
e-ISSN: 2581-2017

Characteristic Estimator of Interval-Censored Binomial Data and Its Application

Bernadhita H. S. Utami[1], Dwi Herinanto[2], Miswan Gumanti[3]
[1,2,3]System Information, STMIK Pringsewu, Lampung, Indonesia

Article Info

Article history:

Received : 03-13-2022
Revised : 06-03-2022
Accepted : 11-13-2022

Keywords:

Binomial distribution;
Interval-censored;
Maximum likelihood estimation;
Survival analysis.

ABSTRACT

This study aims to determine the estimation of interval-censored data with a special distribution, namely the binomial distribution. This research is using quantitative methods, the steps are estimating parameters on the interval-censored binomial distribution using the Maximum Likelihood Estimation method. The second step shows the properties of the estimator on the interval-censored binomial distribution. The last is to determine the parameter estimation of interval-censored data from the binomial distribution in survival analysis and provide an example of research containing interval-censored observations which will then be used as a case study. The results showed that the estimator is a sufficient statistic, meaning that it is unbiased. The case study was conducted using interval-censored data regarding the study of ninety-four breast cancer patients to see which group survived longer (survival value) of the two treatments, namely patients who underwent radiotherapy alone and patients who underwent radiotherapy followed by adjuvant chemotherapy.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v6i1.1847

Corresponding Author:

Bernadhita H. S. Utami,
Department of System Information, STMIK Pringsewu
Email: bernadhita.herindri.s@mail.ugm.ac.id

A. INTRODUCTION

The binomial probability distribution is a discrete probability distribution that is most often used to represent events in everyday life (Zhou et al., 2018). The characteristics of the binomial distribution are that it has only two possible outcomes in an experiment, the random variable is the number of successful events calculated from several trials, and the probability of a successful event remains the same even though the experiment is repeated several times, and each trial is independent of each other, which means the results of one experiment do not affect the results of another experiment (Lehmann and Scheffé, 2012).

In inferential statistics, point estimation is studied, which is a method for determining a single value derived from a sample and used to estimate population parameters (Utami et al., 2021). Regarding parameter estimation, in 1800 Karl Pearson introduced the oldest estimation method called the Moment Method which is a relatively simple method and produces a consistent estimator (Zeng et al., 2016). However, the moment method often produces biased estimators, so to solve this problem Ronald Fisher in 1913 introduced a method called the Maximum Likelihood Estimation method with the principle of maximizing the likelihood function provided that the random sample follows a certain probability distribution (Gupta, B, 2016) (Astapa et al., 2018).

If X is a random variable with a Binomial distribution (n, p) with n known number of trials, then the parameter probability of a successful event p will be estimated based on information about X (Gareth et al., 2013). To obtain an estimator for p with the number of successful events known accurately, the Maximum Likelihood Estimation method is used and we obtained $\hat{p} = \frac{X}{n}$, where X represents the observed success events (Ma, 2010). However, if it is known that there are many trials while the number of successful events is only known to lie in an interval, for example, $[x_1, x_2]$ where x_1 and x_2 are integers between 0 and n, then a slightly different estimate is obtained (Frey and Marrero, 2008).

Several previous studies, introduced the maximum likelihood estimation method to create a semiparametric survival model with censored data on intervals and healing fractions (Zhou et al., 2018). (Szabo et al., 2020) determined the large sample nature of the resulting estimator and evaluated the performance of the finite sample through a simulation study using the maximum likelihood estimator method. For purposes of illustration, the proposed method is applied to the analysis of diabetes conversion data collected from intervention studies of individuals at high risk of developing diabetes. Furthermore, in 2022, (Shen et al., 2022) built the likelihood function and obtained a nonparametric maximum likelihood estimate from the regression parameter using the expectation-maximization algorithm. One of the problems encountered in parameter estimation is the presence of incomplete observations, which in general can be grouped into censored data and truncated data. The incompleteness of the data obtained can be caused by several factors such as limited information, limited resources, or unexpected things. Different circumstances can produce different types of sensors (Singh and Totawattage, 2013). One thing that can be known from the censored interval is the distance (range), which is an interval, which is at the time of the event (Zeng et al., 2016). Thus, through this study, the author will determine the estimation of interval-censored data with a special distribution, namely the binomial distribution. The author will use the maximum likelihood estimation method and determine the characteristics of whether the estimator is an unbiased and minimum variance estimator which is the novelty of this research.

B. LITERATURE REVIEW

Given the random variable Binomial distribution X with parameters n and p, then the probability density function of the random variable X is expressed as

$$b(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, 2, \ldots, n \tag{1}$$

If the number of observations n is known and the random variable X which represents the number of successful events are only known to be in the interval $[x_1, x_2]$ where x_1 and x_2 are integers between 0 and n or expressed as $0 < x_1 < x_2 < n$ then the estimator can be determined by using the Maximum Likelihood Estimation method.

Since X is a random variable from a single observation, the likelihood function obtained is not a multiplication of the probability density function of a random sample $X_1, X_2, \ldots, X_n$ which is expressed as

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i) \tag{2}$$

rather it is expressed as the sum of x_1 to x_2 of the probability density function expressed as

$$L(b) = \sum_{i=x_1}^{x_2} \binom{n}{i} p^i (1-p)^{n-i} \tag{3}$$

Now, review the interval $0 < x_1 \leq x_2 < n$. Notice the order statistic, if the random variables have identical independent distribution $X_1, X_2, \ldots, X_n$ arranged and written as $X_{(1} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ then $X_{(i)}$ is referred to as the i-th order for $i = 1, 2, \ldots, n$ then the cumulative distribution function and probability density function of the r-th order statistic are expressed as follows:

1. Cumulative distribution function

$$F_{(r)}(x) = Pr(X_{(r)} \leq x) \tag{4}$$

$$= \sum_{i=r}^{n} \binom{n}{i} F^i(x)(1-F(x))^{n-i} \tag{5}$$

so that the cumulative distribution function of the 1st order statistic, namely $X_{(1)}$ or also called $X_{(minimum)}$ is expressed as

$$F_{(1)}(x) = \sum_{i=0}^{n} \binom{n}{i} F^i(x)(1-F(x))^{n-i} - (1-F(x))^n = 1-(1-F(x))^n \tag{6}$$

while the statistical cumulative distribution function of the n-th order, namely $X_{(n)}$ or also called $X_{(maximum)}$ is expressed as

$$F_{(n)}(x) = \sum_{i=n}^{n} \binom{n}{i} F^i(x)(1-F(x))^{n-i} = (F(x))^n \tag{7}$$

DOI: https://doi.org/10.30812/varian.v6i1.1847

2. Probability density function The probability density function of the r-th order statistic is defined as follows:

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!}(F(x))^{r-1}(1-F(x))^{n-r}f(x) \tag{8}$$

where $f(x)$ random variable probability density function.
So that the cumulative distribution function of the 1st order statistics, namely $X_{(1)}$ or also called $X_{(minimum)}$ is expressed as

$$f_1(x) = n(1-F(x))^{n-1}f(x) \tag{9}$$

While the statistical probability density function of the n-th order, namely $X_{(n)}$ or also called $X_{(maximum)}$ is expressed as:

$$f_n(x) = n(F(x))^n f(x) \tag{10}$$

C. RESEARCH METHOD

This research is quantitative and the steps taken in this research are as follows:

1. Estimating parameters on the interval-censored binomial distribution using the Maximum Likelihood Estimation method.
2. Shows the properties of the Maximum Likelihood Estimation method estimator on the interval-censored binomial distribution.
3. Studying the parameter estimation of interval-censored data from the binomial distribution in survival analysis.
4. Provide an example of research containing interval-censored observations which will then be used as a case study.
5. Estimating the survival function on interval-censored data using the Maximum Likelihood Estimation method with the help of the R program version 3.3.0.

D. RESULTS AND DISCUSSION

Following the research stages that have been designed, the first step is to estimate the parameters of the interval-censored binomial distribution using the MLE method. Equation (3) can be expressed as follows:

$$L(b) = \sum_{i=x_1}^{x_2}\binom{n}{i}p^i(1-p)^{n-i} - \sum_{i=x_2+1}^{n}\binom{n}{i}p^i(1-p)^{n-i} \tag{11}$$

The last equation is the difference from the form of the cumulative distribution function of statistical order so that each has a probability density function as

$$f_{x_1}(x) = \frac{n!}{((x_1-1)-1)!(n-x_1)!}p^{x_1-1}(1-p)^{n-x_1} \tag{12}$$

and

$$\begin{aligned} f_{x_2}(x) &= \frac{n!}{((x_2+1)-1)!(n-(x_2+1))!}p^{x_2+1}(1-p)^{n-(x_2+1)} \\ &= \frac{n!}{x_2!(n-x_2-1)!}p^{x_2}(1-p)^{n-x_2-1} \end{aligned} \tag{13}$$

Pay attention to the Beta function

$$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \quad ,\alpha>0,\beta>0 \tag{14}$$

and

$$B(k,n-k+1) = \int_0^1 t^{k-1}(1-t)^{n-1}\,dt = \frac{(k-1)!(n-k)!}{n!} \tag{15}$$

By using the relationship between the Binomial Distribution and the Beta Distribution, the likelihood function in equation (3) can be expressed as

$$\begin{aligned} L(b) &= I_p(x_1, n-x_1+1) - I_p(x_2+1, n-x_2) \\ &= \int_0^p \frac{n!}{(x_1-1)!(n-x_1)!}y^{x_1-1}(1-y)^{n-x_1}dy - \int_0^p \frac{n!}{x_2!(n-x_2-1)!}y^{x_2}(1-y)^{n-x_2-1}dy \end{aligned} \tag{16}$$

DOI: https://doi.org/10.30812/varian.v6i1.1847

Differentiation of $L(b)$ with respect to p is

$$\begin{aligned}\frac{dL(b)}{dp} &= \frac{d}{dp}\int_0^p \frac{n!}{(x_1-1)!(n-x_1)!} y^{x_1-1}(1-y)^{n-x_1}dy - \int_0^p \frac{n!}{x_2!(n-x_2-1)!}p^{x_2}(1-p)^{n-x_2-1}dy \\ &= \frac{n!}{(x_1-1)!(n-x_1)!}p^{x_1-1}(1-p)^{n-x_1} - \frac{n!}{x_2!(n-x_2-1)!}p^{x_2}(1-p)^{n-x_2-1}dy\end{aligned} \tag{17}$$

The estimator is obtained by equating $\frac{dL(b)}{dp}$ equal to zero

$$\Leftrightarrow \frac{\hat{p}}{1-\hat{p}} = \left(\frac{x_2!(n-x_2-1)!}{(x_1-1)!(n-x_1)!}\right)^{\frac{1}{x_2-x_1+1}} \tag{18}$$

Note that

$$x_2! = x_2\cdot(x_2-1)\cdot(x_2-2)\ldots(x_1+2)\cdot(x_1+1)\cdot x_1\cdot(x_1-1)! \tag{19}$$

$$(n-x_1)! = (n-x_1)(n-x_1-1)(n-x_1-2)\ldots(n-x_2+2)(n-x_2+1)(n-x_2)(n-x_2-1)! \tag{20}$$

Therefore,

$$\Leftrightarrow \frac{\hat{p}}{1-\hat{p}} = \left(\left(\frac{x_1!}{(n-x_1)!}\right)\left(\frac{x_1!+1}{(n-x_1-1)!}\right)\cdots\left(\frac{x_2!}{(n-x_2)!}\right)\right)^{\frac{1}{x_2-x_1+1}} \tag{21}$$

To facilitate the calculation of the power form in the last equation, logarithmic operations are carried out so that we get

$$\Leftrightarrow \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right) \tag{22}$$

In addition to making the calculation simpler, the last form of the equation is log-odds, which is a log comparison of the probability of a successful event to the probability of a failure event. The log odds form in equation (22) is the mean of the MLE estimator.

Thus, the MLE estimator of the odds of success log is the mean of the MLE odds of success log obtained by assuming the number of successful events for each integer value in the interval $[x_1, x_2]$. The value of the p estimator is obtained by solving equation (22) as follows:

$$\Leftrightarrow \hat{p} = \frac{\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1+\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)} \tag{23}$$

It will be shown that $\hat{p}$ global maximum, in other way that the second derivative of p is negative, considering equation (17),

$$\frac{dL(b)}{dp} = \frac{n!}{(x_1-1)!(n-x_1)!}p^{x_1-1}(1-p)^{n-x_1} - \frac{n!}{x_2!(n-x_2-1)!}p^{x_2}(1-p)^{n-x_2-1} \tag{24}$$

To simplify the differentiation, the natural logarithm function is used as follows:

$$\begin{aligned}\ln\frac{dL(b)}{dp} &= \ln\left(\frac{n!}{(x_1-1)!(n-x_1)!}\right) + (x_1-1)\ln p + (n-x_1)\ln(1-p) \\ &\quad - \ln\left(\frac{n!}{x_2!(n-x_2-1)!}\right) - x_2\ln p - (n-x_2-1)\ln(1-p)\end{aligned} \tag{25}$$

Thus obtained

$$\frac{d}{dp}\left(\ln\frac{L(b)}{dp}\right)=(x_1-x_2-1)\left(\frac{1}{p(1-p)}\right) \tag{26}$$

By substituting the value of $\hat{p}=\dfrac{\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1+\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}$ into p and remembering that $0<x_1\leq x_2<n$

where x_1, x_2, n are integers then

$$(x_1-x_2-1)\left(\frac{1}{\hat{p}(1-\hat{p})}\right)<0$$

Furthermore, we will examine the statistical properties of the interval-censored binomial distribution. Consider the probability density function of the Binomial Distribution with parameters (n,θ):

$$f(x|\theta)=\binom{n}{x}(1-\theta)^n\exp\left(x\log\left(\frac{\theta}{1-\theta}\right)\right) \tag{27}$$

so that $f(x|\theta)$ can be parsed into the form:

$$h(x)=\begin{cases}\binom{n}{x} & ;x=0,1,2,\ldots,n\\ 0 & ;others\end{cases} \tag{28}$$

$$c(\theta)=\begin{cases}(1-\theta)^n & ;0<\theta<1\\ 0 & ;others\end{cases} \tag{29}$$

$$w_1(\theta)=\begin{cases}\log\left(\frac{\theta}{1-\theta}\right) & ;0<\theta<1\\ 0 & ;others\end{cases} \tag{30}$$

$$t_1(\theta)=x \tag{31}$$

Thus, the Binomial Distribution is a class of the exponential family. Based on the theorem which states that an exponential family class has a statistic $T(\underline{X})$, then $T(\underline{X})$ is a sufficient statistic for θ. Therefore,

$$T(\underline{X})=\frac{\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1+\exp\left(\frac{1}{x_2-x_1+1}\left(\sum_{i=x_1}^{x_2}\log\frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)} \tag{32}$$

is a sufficient statistic.

Based on the estimation results for p in equation (23) the first property is obtained, namely that the $\hat{p}$ estimator always lies between $\frac{x_1}{n}$ and $\frac{x_2}{n}$ for $x_1<x_2$. The second property is that the function $\log\left(\frac{p}{1-p}\right)$ is symmetric about $p=0,5$. Furthermore, the value of $\hat{p}$ will be equal to $\frac{x_1+x_2}{2n}$ if and only if $x_1+x_2=n$ or $x_1=x_2$. Note that for many trials n it can be determined all possible intervals $[x_1,x_2]$ in n where $0<x_1\leq x_2<n$. Thus, can be shown the symmetry properties of the $\hat{p}$ estimator for interval-censored binomial data based on the difference between $\hat{p}-\frac{x_1+x_2}{2n}$ (Anderson-Bergman, 2017). The following plots of

DOI: https://doi.org/10.30812/varian.v6i1.1847

the difference of $\hat{p} - \dfrac{x_1 + x_2}{2n}$ to the mean $\dfrac{x_1 + x_2}{2n}$ for the number of trials $n = 5$, $n = 10$, and $n = 20$ are shown in Figure 1 to Figure 3.

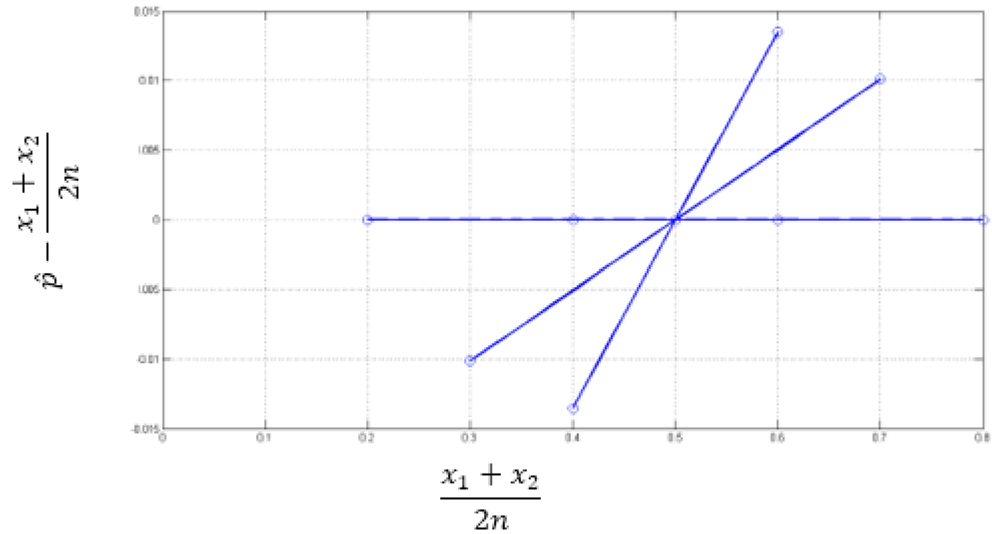

Figure 1. Plot $\hat{p}$ for $n = 5$

For $n = 5$ then there are $\binom{5}{2} = 10$ possible intervals $[x_1, x_2]$ that can be made. Figure 1 shows that the maximum difference exists when the average value is 0.6, which is 0.01351179. This value is obtained during the interval $[2, 4]$. It appears that the plot is symmetrical to the mean of 0.5.

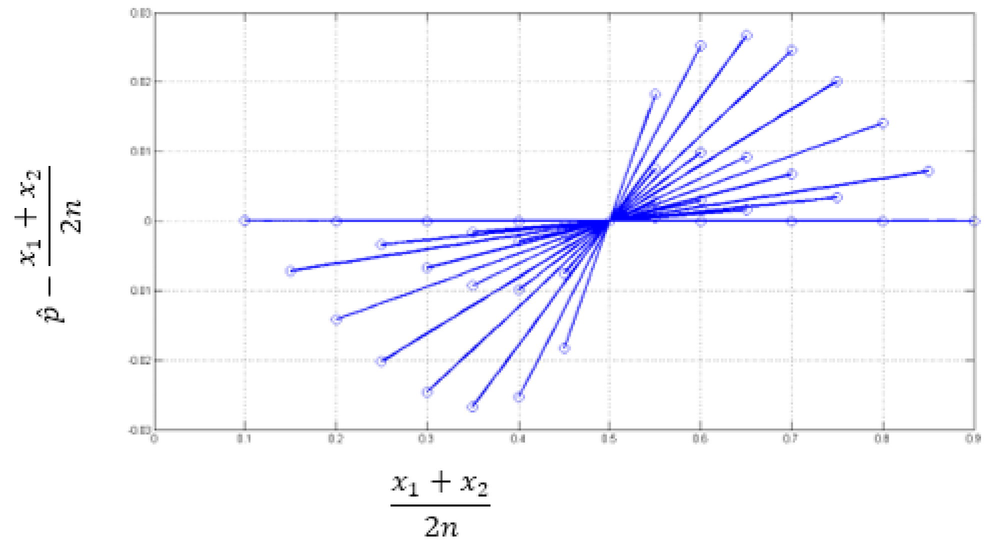

Figure 2. Plot $\hat{p}$ for $n = 10$

Figure 2 shows a plot for $n = 10$ so there are $\binom{10}{2} = 45$ possible intervals $[x_1, x_2]$ that can be made. From the plot, it is shown that the maximum difference exists when the average is 0.65, which is 0.026661. This value is obtained during the interval $[4, 9]$. It appears that the plot is symmetrical to the mean of 0.5.

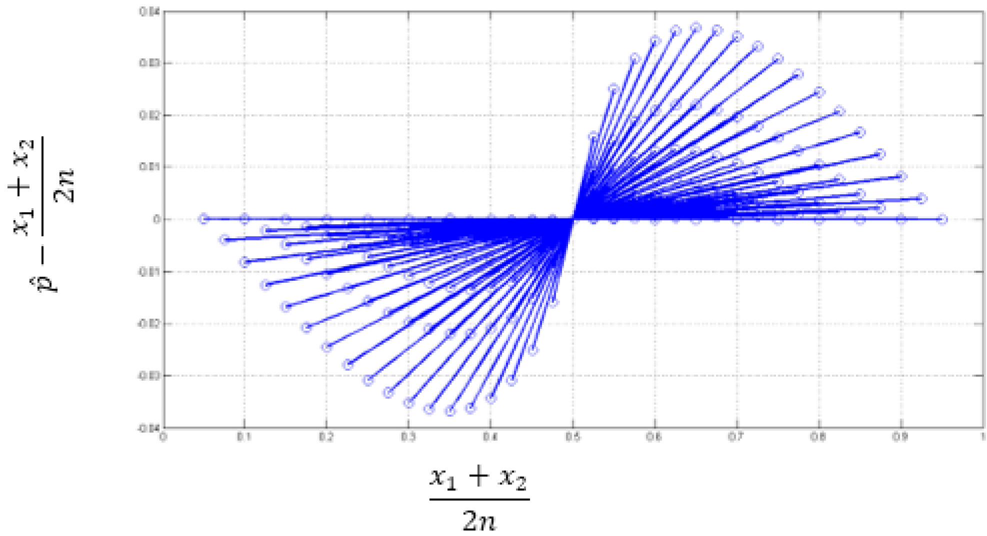

Figure 3. Plot $\hat{p}$ for $n = 20$

From Figure 1 to Figure 3 it can be concluded that for many experiments with increasing n the plot shows an upward concave pattern for probability values 0 to 0.5 and shows a downward concave pattern for probability values 0.5 to 1.

The following is an application of the interval-censored Binomial distribution in survival analysis. Interval censored observations are represented in the form of two times, namely T_L which states the censored time from the left, and T_R which states the censored time from the right (Wang et al., 2016). The likelihood function for interval-censored observations is expressed by

$$L = \prod_{i=1}^{n} S(T_{Li}) - S(T_{Ri}) \tag{33}$$

Pay attention to the cumulative distribution function of random variables with Binomial distribution

$$F(t) = \sum_{k=0}^{t} \binom{n}{k} p^k (1-p)^k \tag{34}$$

so the survival function is

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \sum_{k=0}^{t} \binom{n}{k} p^k (1-p)^k \end{aligned} \tag{35}$$

Thus the likelihood function of the censored interval is as follows:

$$L = \prod_{i=1}^{n} \left(\sum_{k=0}^{tRi} \binom{n}{k} p^k (1-p)^k \right) - \left(\sum_{k=0}^{tLi} \binom{n}{k} p^k (1-p)^k \right) \tag{36}$$

From this likelihood function, an estimator will be obtained by substituting the time interval into the equation. Henceforth, estimates are obtained by computational analysis which is discussed in the next section. The data used as an example in this study is a retrospective study to compare patients with early-stage breast cancer who received radiation therapy and adjuvant chemotherapy with patients who received radiation therapy alone on the cosmetic effects of each treatment (Shen and Cook, 2015). Excisional biopsy (removal of the entire tumor) followed by radiation is an increasingly practiced alternative to mastectomy (removal of the breast). Adjuvant chemotherapy in the form of giving anti-cancer drugs after surgery, and radiotherapy will suppress cancer cells so that the patient does not relapse within a certain time (Chen et al., 2012). The long-term effects of adjuvant chemotherapy in patients previously undergoing radiotherapy are uncertain. The purpose of this analysis was to compare patients who underwent chemotherapy after undergoing radiotherapy with patients who only underwent radiotherapy in terms of the degree of cosmetic deterioration. Patients undergo therapy every four to six months with a steadily decreasing frequency (Singh and Totawattage, 2013).

The doctor recorded the patient's cosmetic appearance regarding breast edema, telangiectasia, breast retraction, and all cosmetic results. Therefore, a comparison will be made between the two 23 treatments at the time of cosmetic damage. The object of research in this study was breast cancer patients who were treated at the Joint Center for Radiation Therapy in Boston from 1976 to 1980. This breast cancer data has a total of 94 observations, 56 of which are interval-censored and 38 are right-censored data.

Table 1. Estimated survival value of patients undergoing radiotherapy

Time (week)	Y	d	Survival value
4.5	46	2	0.954
6.5	44	2	0.920
7.5	42	4	0.832
11.5	38	3	0.761
15.5	35	0	0.761
17.5	35	0	0.761
24.5	35	4	0.668
25.5	31	0	0.668
33.5	31	4	0.587
34.5	27	0	0.587
36.5	27	0	0.586
39.0	27	6	0.466
42.0	21	0	0.466
47.0	21	2	0

Based on the results above, it can be interpreted that at time 0 the survival value is 1 with the number of patients at risk being 46 people, but no event occurred (n event = 0) so the number of patients at risk the next time is still 46 people. Furthermore, at the 4th time, the survival value decreased to 0.954 with 46 patients at risk and 2 events occurred. At the 6.5th time, the survival value became 0.920 with 44 patients at risk and 2. Furthermore, at the time of 11.5, the survival value became 0.761 with 38 patients at risk. The survival rate of 0.761 occurred until the 17.5th time with 35 patients at risk. The survival value reached 0 for the 47-th time with the remaining 21 patients at risk.

The following is the result of the estimated survival value for patients who underwent radiotherapy and continued with chemotherapy

Table 2. Estimated survival value of patients undergoing radiotherapy followed by adjuvant chemotherapy

Time (week)	Y	d	Survival value
4.5	48	2	0.9567
6.5	46	2	0.9134
11.5	44	3	0.8442
12.5	41	0	0.8442
16.5	41	7	0.6987
18.5	34	7	0.5580
19.5	27	6	0.4420
21.5	21	0	0.4420
22.5	21	0	0.4420
23.5	21	0	0.4420
24.5	21	5	0.3421
30.5	16	3	0.2714
31.5	13	0	0.2714
33.5	13	0	0.2711
34.5	13	0	0.2711
35.5	13	8	0.1105
45.5	5	3	0.0552
54.0	2	2	0

Based on the results above, it can be interpreted that at time 0 the survival value is 1 with the number of patients at risk being 48 people, but no event occurred (n event = 0) so the number of patients at risk the next time is still 48 people. Furthermore, at the 4.5th time, the survival value decreased to 0.9567 with 48 patients at risk and 2 events occurred. Thus, the number of patients at risk at the 6.5th time was 46 people with a survival value of 0.9134. In the 16.5 to the 23.5th-time interval, 21 patients had a survival

DOI: https://doi.org/10.30812/varian.v6i1.1847

probability value of 0.442. The survival value reached 0 at the 54th time with the remaining 2 patients at risk. To make it easier to make a comparison between radiotherapy treatment with a combination of radiotherapy and chemotherapy, a survival function graph is made in Figure 4 as follows:

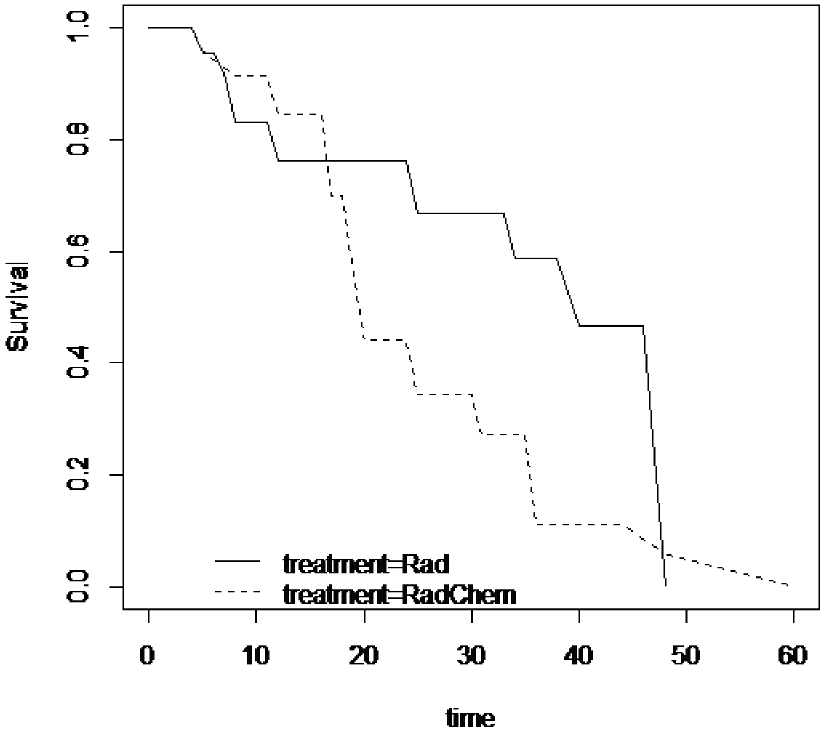

Figure 4. Graph of Radiotherapy Survival Function and Combination Radiotherapy-Chemotherapy

From Figure 4, the straight line represents the graph of the survival function for patients who underwent radiotherapy treatment alone, while the dotted line represents the graph of survival function who underwent radiotherapy and continued chemotherapy. The survival value at the time of 4.5 in the two treatments was not much different, ranging from 0.95. Differences began to appear at the 11.5 to 24.5 time with a survival value of 0.761 for radiotherapy treatment while at the same time interval for radiotherapy-chemotherapy treatment, the survival value decreased from 0.8442 to 0.6957. Graphically, patients who underwent a combination of radiotherapy and chemotherapy survived longer. This is shown at the 60th week of the patient's probability of surviving reaching 0. While the probability of patients undergoing radiotherapy reaches 0 at the 47th week.

E. CONCLUSION AND SUGGESTION

From the estimation results using the Maximum Likelihood Estimation method, the estimator

$$\hat{p} = \frac{\exp\left(\frac{1}{x_2 - x_1 + 1}\left(\sum_{i=x_1}^{x_2} \log \frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}{1 + \exp\left(\frac{1}{x_2 - x_1 + 1}\left(\sum_{i=x_1}^{x_2} \log \frac{\frac{i}{n}}{1-\frac{i}{n}}\right)\right)}; 0 < x_1 \leq x_2 < n$$

It has been shown that the interval-censored Binomial distribution comes from the Binomial Distribution family which is a class of the exponential family so the estimator is statistically sufficient. Based on a case study using data from breast cancer patients, the survival function plot showed that patients who underwent a combination of radiotherapy and adjuvant chemotherapy survived longer from damage to the surface of the breast up to 60 weeks than patients who underwent radiotherapy alone. Further research can be carried out to estimate the interval-censored data from other special discrete distributions such as the Poisson distribution. In addition, research can be continued on the latest data that describes the hazard of a group of patients with diabetes or other cancers.

ACKNOWLEDGEMENT

Gratefully acknowledges to the financial support of the STMIK Pringsewu.

REFERENCES

Anderson-Bergman, C. (2017). icenreg: Regression models for interval censored data in r. *Journal of Statistical Software*, 81(12):123.

Astapa, I. G. A., Suwardika, G., and Suniantara, I. K. P. (2018). Analisis data panel pada kinerja reksadana saham. *Jurnal Varian*, 1(2):59–69.

Chen, D.-G. D., Sun, J., and Peace, K. E. (2012). *Interval-censored time-to-event data: methods and applications*. CRC Press.

Frey, J. and Marrero, O. (2008). A surprising mle for interval-censored binomial data. *The American Statistician*, 62(2):135–137.

Gareth, J., Daniela, W., Trevor, H., and Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Spinger.

Gupta, B (2016). *Introduction to Basic Statistics*. Michigan: University of Michigan.

Lehmann, E. L. and Scheffé, H. (2012). Completeness, similar regions, and unbiased estimation-part i. In *Selected works of EL Lehmann*, pages 233–268. Springer.

Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica*, pages 1165–1181.

Shen, H. and Cook, R. J. (2015). Analysis of interval-censored recurrent event processes subject to resolution. *Biometrical Journal*, 57(5):725–742.

Shen, P.-s., Peng, Y., Chen, H.-J., and Chen, C.-M. (2022). Maximum likelihood estimation for length-biased and interval-censored data with a nonsusceptible fraction. *Lifetime Data Analysis*, 28(1):68–88.

Singh, R. S. and Totawattage, D. P. (2013). The statistical analysis of interval-censored failure time data with applications. *Open Journal of Statistics*, 3(2):155–166.

Szabo, Z., Liu, X., and Xiang, L. (2020). Semiparametric sieve maximum likelihood estimation for accelerated hazards model with interval-censored data. *Journal of Statistical Planning and Inference*, 205:175–192.

Utami, B. H. S., Irawan, A., Gumanti, M., and Primajati, G. (2021). Hausman and taylor estimator analysis on the linear data panel model. *Jurnal Varian*, 5(1):81–88.

Wang, L., McMahan, C. S., Hudgens, M. G., and Qureshi, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*, 72(1):222–231.

Zeng, D., Mao, L., and Lin, D. (2016). Maximum likelihood estimation for semiparametric transformation models with interval-censored data. *Biometrika*, 103(2):253–271.

Zhou, J., Zhang, J., and Lu, W. (2018). Computationally efficient estimation for the generalized odds rate mixture cure model with interval-censored data. *Journal of Computational and Graphical Statistics*, 27(1):48–58.